

EXHIBIT 14

I am writing to appeal the decision of the Committee on Score Validity that it could not certify the validity of my passing level results on the December 2007 Step 2 CK exam as outlined in the letter dated February 17, 2010. I appeal on the basis that (1) the Policies and Procedures Regarding Indeterminate Scores were not followed; and (2) the decision of the Committee on Score Validity was contrary to the weight of the evidence.

I. PROCEDURAL BACKGROUND AND COMMITTEE ON SCORE VALIDITY FINDINGS:

In a letter dated September 15, 2009, Susan Deitch of the Office of the USMLE Secretariat (on page 2 of 5) she stated that *"...an analysis of your performance on the December 31, 2007 Step 2 CK, along with your acknowledged presence at Optima, raises concerns about the validity of the passing level scores reported to you for that examination."* Ms. Deitch cited two reasons for the decision; an "analysis" of my performance and my attendance at the Optima review course.

The letter went on to describe that certain percentages of test items that appeared on my Step 2 CK form *"may have been subject to unauthorized reproduction and dissemination through Optima prior to December 2007."* It is not fair or reasonable that the Committee could base its decision to invalidate my scores on a mere suspicion that certain test questions MAY have been exposed.

The letter also contains percentages of questions that were exposed and those that were not exposed and contains data on the length of time it took for me to answer each subset of questions on average.

In December 2009, I was given an opportunity to defend my position against these allegations. I took this opportunity and am confident that a careful and thoughtful review of the Transcript of Board Meeting on Irregular Behavior from December 16, 2009, will reveal that I clearly defined accurate, clear, and plausible reasons for any discrepancies questioned.

On February 17, 2010, a letter was written to me stating that *"Following careful consideration of all of the available information, the Committee determined that it could not certify the validity of your passing level results on the December 2007 Step 2 CK."*

The letter stated that the Committee reviewed the following items:

- 1- I attended Optima before my Step 2 CK in December 2007.
- 2- I passed my Step 2 CK on my sixth attempt.
- 3- Percentages of items correct in exposed verses unexposed question subsets.
- 4- Time spent in exposed verses unexposed question subsets.
- 5- Information I stated at my hearing:
 - I was at Optima for less than 5 weeks, after having done multiple review courses and study question banks.
 - I benefited from the intense study environment and study groups in Totowa, and used the test bank minimally.
 - I missed passing on my fifth attempt by one point, and kept focus on my weakest and only left starred subject, Obstetrics and Gynecology
 - I rushed at the end of blocks and that may have affected times.

The reasons given to me for not validating were as follows:

- 1- Even though my *"performance on the subset of items for which there is no evidence of exposure (68% of the total number of scored items) was quite close to a level that would be considered*

"passing" for that subset of items" ... "the Committee expressed concerns about using performance on a subset of items (non-exposed) as a basis for predicting performance on the entire examination."

- 2- "The Committee considered the reasonable possibility that familiarity with exposed items may enable an individual to respond, on average, more quickly to such items and produce a time advantage for the individual in considering and responding to the non-exposed items."
- 3- Given the differences in two subsets, the Committee found it could not presume validity based on my non-exposed percentages.

The following paragraphs will highlight the evidence and information supporting the only fair conclusion of my case, that my test scores should be validated.

II: MY TEST SCORES SHOULD BE VALIDATED BECAUSE THE COMMITTEE ON SCORE VALIDITY DID NOT FOLLOW THE POLICIES AND PROCEDURES REGARDING INDETERMINATE SCORES AND THE WEIGHT OF THE EVIDENCE IS IN FAVOR OF VALIDATING MY SCORE:

A. The Policies and Procedures Regarding Indeterminate Scores were NOT followed.

1. Policies and Procedures Violated:

The following excerpts from the USMLE Policies and Procedures Regarding Indeterminate Scores are relevant to my appeal:

Section A. Policies, policy #1 states "Aberrancy(ies) in performance for which there is no reasonable and satisfactory explanation result in passing scores being classified as indeterminate." Policy #2 states "Statistical procedures will be applied routinely, as well as in response to particular information, to identify scores that may be subsequently classified as indeterminate."

Section B. Procedures, procedure #1 states "These procedures are applicable to instances in which: a) the results of appropriate statistical analyses identify an aberrancy(ies) in performance, i.e., indicate that a score does not or may not represent a reasonable assessment of an examinee's knowledge or competence sampled by the examination. Such statistical analyses include, but are not limited to, analyses which (i) indicate that the pattern of scores for a given examinee is markedly nonuniform and one or more section scores for the examinee is below the passing level; (ii) indicate that the current scores for a given examinee show an unexpectedly large increase over the examinee's most recent prior scores on the same Step;"

Section B. Procedures, procedure #2 describes areas in which staff will review any additional information available from examination records that may be helpful in explaining the aberrancy(ies), such as:

- a) ...in the context of nonuniformity of performance, such further investigation may include, e.g., review of responses to determine whether large numbers of questions were omitted or answered randomly; ...; and review of any information obtained before or after the examination that might be relevant to his or her performance on the test.
- b) ... in the context of an unexpectedly large increase over the examinee's most recent prior score, such further investigation might include ...review of the examinee's record to ascertain whether the examination was taken on multiple prior occasions, and if so, whether the examinee's score on one or more earlier attempts was higher than the examinee's score on the most recent previous examination.
- c) Staff will review the results of such investigation to determine whether a clearly reasonable and satisfactory explanation for the results of the statistical analyses has been obtained.

Section B. Procedures, procedure #5 talks about how an examinee will be notified if there is not enough evidence to validate a score and then the score gets turned over to the Committee on Score Validity. It continues to state "Staff will notify the examinee of the withholding of scores and will provide the examinee with a description of any statistical analyses employed." Procedure 6 states, "All pertinent information, including a description of any statistical analyses employed and any explanation of other information that the examinee may provide, will be presented to the Committee on Score Validity which is authorized to make a determination regarding the validity of the scores in question."

2) *How the USMLE Policies and Procedures Were Violated in My Case*

The above procedures and policies clearly state that the beginning of any process to determine if a score is not a valid measure of an examinee's competence is a statistical analysis of the exam itself and the responses given. In the letter from Ms. Deitch, dated September 15, 2009, it was stated that evidence showed that 32% of my exam may have been "exposed" prior to my taking the exam, that I scored higher on this subset than the "non-exposed" subset, and that I was faster in the "exposed" subset compared to the "non-exposed".

I raised the validity of these finding with Ms. Janet Carson in a telephone conversation in late November 2009. I requested a stratification of the findings and contact information of the statistician who did the analysis, to obtain an official report from him. During this conversation, she told me that the analysis for which I sought stratification was NOT an analysis, but an OBSERVATION. Therefore an analysis was not actually performed in my case. I brought this point up in my hearing (seen on page 16, line 11 onward and page 17, lines 15 onward), and there was no rebuttal to this statement by either the Committee or Ms. Carson. By not rebutting my challenge, they in essence agreed that this was an observation, and that the numbers were observed by NBME employees whose knowledge, qualifications, and background were not reviewed to see if they could make a professional, unbiased evaluation of the dataset. **By this not being a statistical analysis, there is a violation of Policy 1.**

I had also asked for a stratification of the subsets to verify the uniformity among the question types. I described my proficiency in subjects such as Psychiatry (I hold a Bachelor of Science in Psychology), as well as a weakness of knowledge in subjects such as Infectious Disease and Obstetrics and Gynecology. If there was a true analysis, these data sets would have been given to me or to the Committee when I defended myself with this logic. **By not answering these requests and not providing specific data to me or to the Committee, there is a violation of Procedure 2c in that they did not "review the results of such investigation to determine whether a clearly reasonable and satisfactory explanation for the results of the statistical analyses has been obtained."** In fact, I was **denied any discovery** to specific material pertaining to this observation.

I had also explained why certain questions took me a long time answer while others took less time. One question took me four minutes to answer. I answered questions at the end of blocks quickly because I was running out of time and then went back to read them. These explanations were not considered. By not looking into this logic and analyzing the time spent when I was closer to the end of a block, there is a **violation of Procedure 2a in that they did not adhere to the concept that "in the context of nonuniformity of performance, such further investigation may include, e.g., review of responses to determine whether large numbers of questions were omitted or answered randomly"**. Further investigation would have shown a true and reasonable cause for aberrancies in time and items being answered irrelevant of a subset.

By not giving me a clear description of what the observations meant, what they showed, which questions they referred to, and how they matched up against all other test takers, there is a clear violation of Procedure 5 and 6 which state staff "will provide the examinee with a description of any statistical analyses employed" and "All pertinent information, including a description of any statistical analyses employed and any explanation of other information that the examinee may provide, will be presented to the Committee on Score Validity" My points were "pertinent" and neither I nor the Committee was given a more in depth understanding of what the numbers really showed or compared, even after I had requested it weeks prior to the hearing and again during the hearing.

Policy 1 states that any analysis will be done routinely. This was not a routine analysis of scores. It was based on a court case against Optima University, one which is still pending six months AFTER my hearing. In fact, when I took the exam in 2007, I received my score even with a history of prior attempts and with a jump in my score. If the score itself had an aberrancy irrelevant to attendance at Optima University, then an analysis would have been warranted there. **According to the policy, this would only be questioned if a "routine analysis" was done**, not one which evolved from an open litigation against a review course. **This is a further violation of Policy 1.**

Procedure #1 states the procedures are applicable when the results of an "appropriate statistical analyses" identify aberrancy in performance, i.e., indicate that a score does not or may not represent a reasonable assessment of an examinee's knowledge or competence sampled by the examination. Ms. Janet Carson clearly stated that this was not a statistical analysis, but an observation. Two areas of concern to the board were the nonuniform scoring, and that I took the exam multiple times. With regard to the former, I have already established that not all necessary data was released for a proper analysis, and that there are many factors that were not further investigated which could have caused aberrancy. With regard to the latter, I made it clear in the hearing (Transcript page 10, lines 8 onward) that I took the exam multiple times, increasing my score each time. In the attempt before my pass, I had failed by only one point. This shows a pattern of continued improvement and that I was already bordering at the level to pass during my last attempt. I also stated on page 42, lines 1 to 6, that I took the exam multiple times when I was not ready for it because I had paid for it. **If both of these points were that great of concern, NBME/USMLE should have called for an analysis BEFORE my score was reported to me, irrelevant of attendance at Optima University.**

In summary, the committee on score validity did not follow the proper policies and procedures. The policy of the USMLE calls for a statistical analysis of my exam to validate scores. NBME officials stated the findings were not a statistical analysis, but an **observation** made by NBME employees who did not have a statistical proficiency. The observations must **NOT** be used in the decision making of this committee. Procedures also call for further investigation when there are possible causes for aberrancies. I raised many points to cause a need for further investigation, and the NBME/USMLE did not investigate more to clarify these queries. In addition, procedures call for a description of any statistical analyses and pertinent information to be disclosed. Even after multiple requests, complete descriptions of the pertinent information were not released for review. Finally, Policy 1 calls for review if a routine analysis finds aberrancy. This was not a routine analysis, but one that was started based on unproven allegations against a review course. The analysis is illegitimate as the root is still unfounded.

B. The Decision of the Committee on Score Validity was Contrary to the Weight of the Evidence.

The reasons given to me for not validating were as follows:

- 1- *the Committee expressed concerns about using performance on a subset of items as a basis for predicting performance on the entire exam*
- 2- *The Committee considered a reasonable possibility that familiarity with exposed items may enable a person to respond, on average, more quickly to such items and gain a time advantage to responding to non-exposed items.*
- 3- *Given differences in two subsets, the Committee could not presume validity based on non-exposed percentages.*

The Committee on Score Validity reviewed just 4 factors from my forty-five minute hearing as stated in their letter on February 17, 2010:

- 1- I was at Optima for less than 5 weeks, after having done multiple review courses and study question banks.
- 2- I benefited from the intense study environment and study groups in Totowa, and used the test bank minimally.
- 3- I missed passing on my fifth attempt by one point, and kept focus on my weakest and only left starred subject, Obstetrics and Gynecology
- 4- I rushed at the end of blocks and that may have affected times.

Throughout my hearing, I expressed concern about the findings in the observation held by the NBME/USMLE regarding exposed items on my exam. Janet Carson stated that they knew what Forms of the USMLE Optima University allegedly had access to and reproduced. They also claimed that they knew it was prior to my exam in 2007. On transcript page 34, lines 6 to 10, Janet Carson states *"I referred to the fact that analyses for you were based on the forms of the exam exposed before you tested. I did not identify the point in time in which we learned of the exposures."* This statement was in rebuttal to me stating that the NBME/USMLE should have pulled the exam forms before I tested if they knew they were exposed. One can see that the defensive statement shows that they knew about the exposure AFTER my exam. There were only a few instances when Optima materials were seized: all of them in May 2008 and after in Tennessee. I stated numerous times that the review question bank in Optima University was updated in March 2008. That was almost three months **after** I took my exam. If an "analysis" or comparison of my exam was done based on a bank obtained after that date, there could be questions that were never in the Optima bank at my time of being a student. However, NBME/USMLE found it acceptable to **PRESUME** I had access to these questions and used this presumption against me. Again, a stratification of the subsets could start a process by which to test this theory. Dr. Whelan stated *"...the fact that we know, we have – the confiscated files represent actual test data. So we know its there."* (Transcript page 32, lines 4 to 6). If evidence exists, then there is a creation date of the file, and if created after my exam date, then we know I had no access. **NBME/USMLE refused to do this further investigation. They instead presume my guilt of exposure.**

As I stated earlier, I asked for a stratification of the observation at multiple times. There is a firm correlation between my areas of proficiency and my areas of weakness which affect my percentages of right and wrong. There is a direct correlation if they fell under the exposed or non-exposed subsets. This also affects the speed at which I would do a question. Speed is not determined just by previous exposure, but also by the strength of knowledge and recognizing key words. **They did not even consider this argument, notwithstanding that I provided clear support based on my test taking strategy and coursework strength.**

I stated that Optima University was focused on Step 1 students throughout 2007. During my short time period there, the course had less than five Step 2 students since its inception in March 2007. There was a very minimal bank and the Step 2 students gained more from self study and group discussion (started by students and not part of his program) in a hard core environment other than any course material. We did more USMLEWorld questions than the Optima bank. Students also did NBME assessment exams. Students used a mix of the Optima bank, other question banks, review groups, lectures and assessment exams and books throughout the course. All factored into a student's performance. **This point was not considered relevant.**

I stated that I was one point from passing and that I had successively increased my score. The Committee has given more weight to my attempts and the presumed access to material than seeing my trend towards passing.

I stated that I was being compared to an average score of a control group of over 1000 students. They did not call me an outlier in any form. In fact, they did not address the points I made on Transcript page 18, line 1 onward, in which I asked about those students who had higher aberrancies of percentages than me. Averages compare high scores and low scores. There was no clarification of my status in this observation. There was also no representation of the data sets in relation to my time on each question. The evidence presented against me was an average of time spent on each question, which needed further review. **They refused to show that this was a reasonable consideration.**

I stated that my focus was on Obstetrics and Gynecology because of the star on the left in my previous two exams, March and July of the same year as my passing score, 2007. This warranted a stratification of the exposed verses non-exposed to see how many questions fell in each subset. It also warranted a comparison to past performance to see if there was a uniform increase in productivity verses a profound spike in performance in Obstetrics and Gynecology. **The Committee didn't consider this possibility.**

I stated that the observations done were by NBME employees, and not by licensed or professional statisticians. They didn't consider that all relative angles were not being considered. They accepted the only means of aberration was that there existed an alleged subset of exposed verses non-exposed questions, and this candidate did better in one than the other. They didn't question stratification, comparison of the stratification to past experience, and what time during the test these questions were seen (beginning, middle, or end of a block). **The observations were flawed in both procedure and content, but this was not considered by the Committee.**

During the course of the hearing, other areas that were irrelevant to my Step 2 CK were brought into conversation by members present with the Committee. On page 42 of the hearing Transcript, Mr. Seeling brings up my Step 1 history. During different areas, others bring up my employment well after my exam date. They also bring up my transfers of medical schools, and my not taking Step 3 until 2009. **NONE of this has to do with my Step 2 CK in December 2007, but the negative connotations may have biased certain discussions which affected the fairness of a decision.**

I talked about the discrepancies of the numbers given. I stated that it was claimed that I did eight more questions correctly than the control group in the exposed group, and got 17 fewer questions correct in the non-exposed group. At the hearing, Dr. Grande tried to state that you must take both aspects of the numbers which go against the candidate. This is unfair in practice as you should assume the same direction. If you give the benefit to the control group when they do better, then you should give the benefit to the candidate, especially when you will not stratify the question base. All in all, we are talking about me getting **1.2 more questions correct per block**, compared to the control group, which in itself is an average. **The Committee refused to accept this as needing clarification by a true statistician.**

The understood three digit score is based on a 288 question subset. Of all administered question in Step 2, 288 questions are actually used in scoring the exam. The number of questions right from this 288 is what your three digit score on the USMLE becomes. At the time of my exam, 184 was the passing score. That is getting approximately sixty-four percent (64%) of the questions correct. The observation stated that I received sixty-six (66%) on my non-exposed items. We should be able to accept that knowledge doesn't change toward the negative if two subsets are equal in nature. There is no stratification supplied. Sixty-six percent of 288 (a number confirmed to be the number of questions used in scoring my exam by Janet Carson in an email to me dated November 11, 2009) is **190.08**. **This would be a passing score. The Committee states they cannot PRESUME the same level of achievement in the exposed subset. This is NOT taking the weight of my evidence into their decision.**

With that said, let me bring into proof three scenarios:

Student	Exam Date	% Exposed	% correct exposed	% correct unexposed	Time spent exposed	Time spent unexposed	Committee Score Validity Decision
A	12/31/07	32%	84%	66%	59 sec	73 sec	INVALID
B	08/19/08	30%	88%	67%	56 sec	80 sec	Valid
C	12/11/08	29%	81%	68%	60 sec	71 sec	Valid

We have here three students from Optima. I know which student took what exam on which day, as a log was kept in New Jersey. All three students went before the Committee of Score Validity. Student B is a Step 2 student and Student C is a Step 1 student. This information can be verified by NBME based on test date and appearance before committee.

When applying the percentages to the number 288, we get the following:

Student	Number of Questions Exposed Correct	Number of Questions Non-exposed Correct	Total Correct
A	77	129	206
B	76	135	211
C	68	139	207

According to the decision letter against me, *the Committee states they could not presumably validate my score based on my performance on the non-exposed subset because of the differences in scoring between my exposed and non-exposed subsets.* My percentages were very similar to Students B and C. They have similar differences in scoring. Also, they both received a higher final score than me on the USMLE. Student B was even quicker than me on her exposed questions and took longer on her unexposed. She was also at Optima for well over six months, stayed after the FBI raid, and was the sister of the Optima owner's right hand man. Student C's time is almost identical as mine, and she was at Optima for well over 6 months, and was in both New Jersey and Tennessee. Ironically, Student B was seen the same day as me by the Committee, and therefore by the same Committee members. The number of questions which differentiate these three students is minute, yet the weight of the final decision is tremendous.

Neither Student B nor C has a passing score based just on non-exposed questions correct, so there must have been a PRESUMED level of achievement in the exposed Question set based on performance in the non-exposed. **There is clearly a discrepancy in how the Committee handled my case verses theirs, and a level of bias and discrepancy which could have resulted from the negative connotations of my work history and Step 1 history.**

Clearly, the decision of the Committee on Score Validity was contrary to the weight of the evidence. There was a double standard when the Committee heard my case versus the cases of other students. They took a harder stance against me and did not give proper review or consideration to the

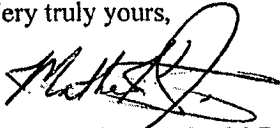
evidence presented before them by me. Instead, the NBME/USMLE presumes that I had access to all forms that Optima allegedly had access to, regardless of the fact that he did updates months after my exam. They also presumed that I had access to all forms that Optima allegedly had access to regardless of the fact they did not get evidence of confiscated files until after May 2008. The Committee did not take into consideration that the observation was just that, an unprofessional observation and NOT a professional statistical analysis. A professional statistician was never utilized for an evaluation. The specifics of the data sets, as well as stratified question types directly affected performance measures and time validity. They also did not consider that the control group was an average and that my scores were not differentiated as to where they lie in that average. The Committee refused to acknowledge that my performance spike was more so in Obstetrics and Gynecology, and that improvement alone would give me a pass from my past exams, as well directly affect performance in exposed verses non-exposed items.

Most importantly, the committee clearly stated in their decision letter that they could NOT presume performance in my case, but obviously did it for two other students with almost identical numbers. There was not a fair and equal hearing afforded to me compared to other students who were at Optima longer than me and who had had access and exposure to the complete question banks (after March 2008) which were seized in Tennessee.

III: CONCLUSION:

Upon review of this letter, the transcripts of my hearing and all documents and materials referenced in my prior correspondence, you will have the evidence to support that (1) the Policies and Procedures Regarding Indeterminate Scores were not followed; and (2) the decision of the Committee on Score Validity was contrary to the weight of the evidence. As such, **I respectfully submit that the only fair and reasonable finding is that my USMLE Step 2 CK examination scores should be validated, and reported to all requesting institutions.** Thank you for your consideration. I look forward to hearing of your decision.

Very truly yours,



Mathew Thomas, Jr., M.D., M.H.S.A.
Submitted via email: June 14, 2010